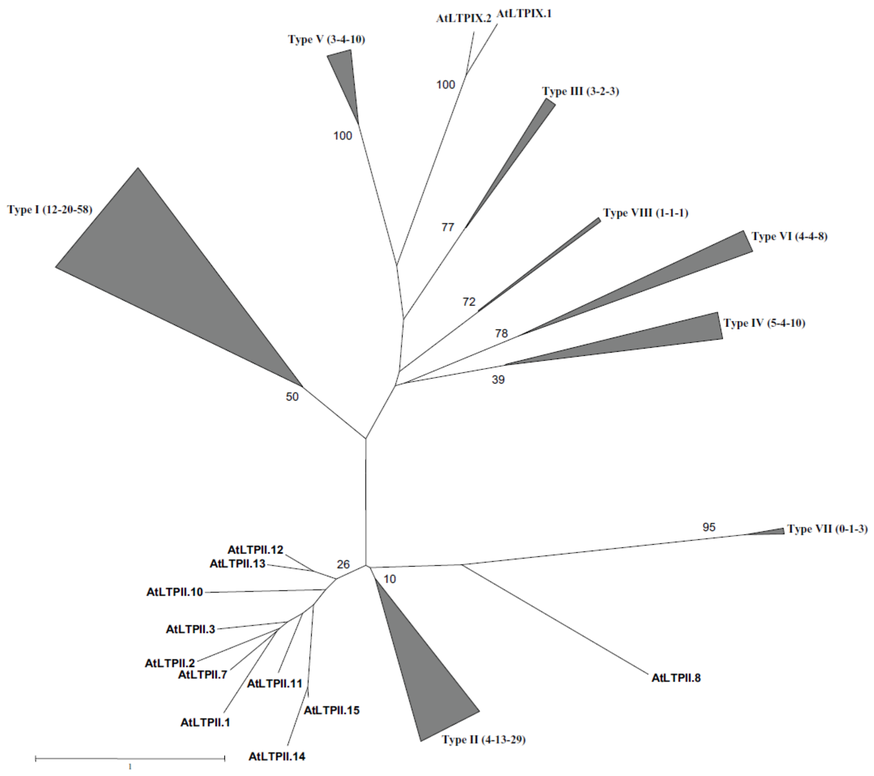
Towards a multi-scale and formalized representation of protein sequence-structure-function relationships : the nsLTP family as a case of study

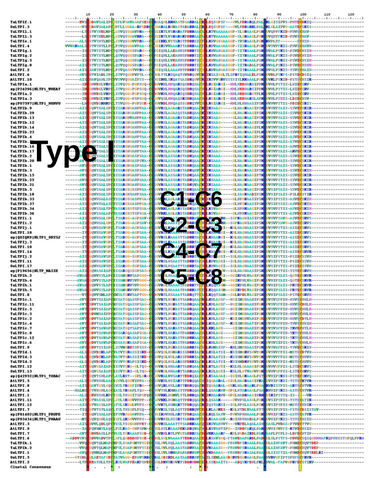
Cécile Fleury[1], Marie-Françoise Gautier[1], Pierre Larmande[1], Sabine Pérès[2], Frédéric de Lamotte[1], Franck Molina[2], Manuel Ruiz[1]

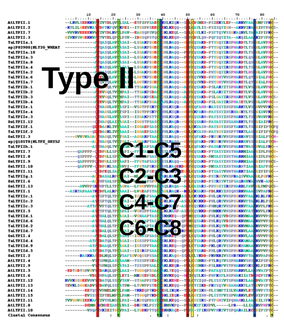
The plant non-specific Lipid Transfer Proteins

- small proteins (6.5-10.5 kDa)
- 8 cysteine motif backbone = C-Xn-C-Xn-CC-Xn-CXC-Xn-C-Xn-C
- nearly 500 different nsLTPs from around 100 plant species

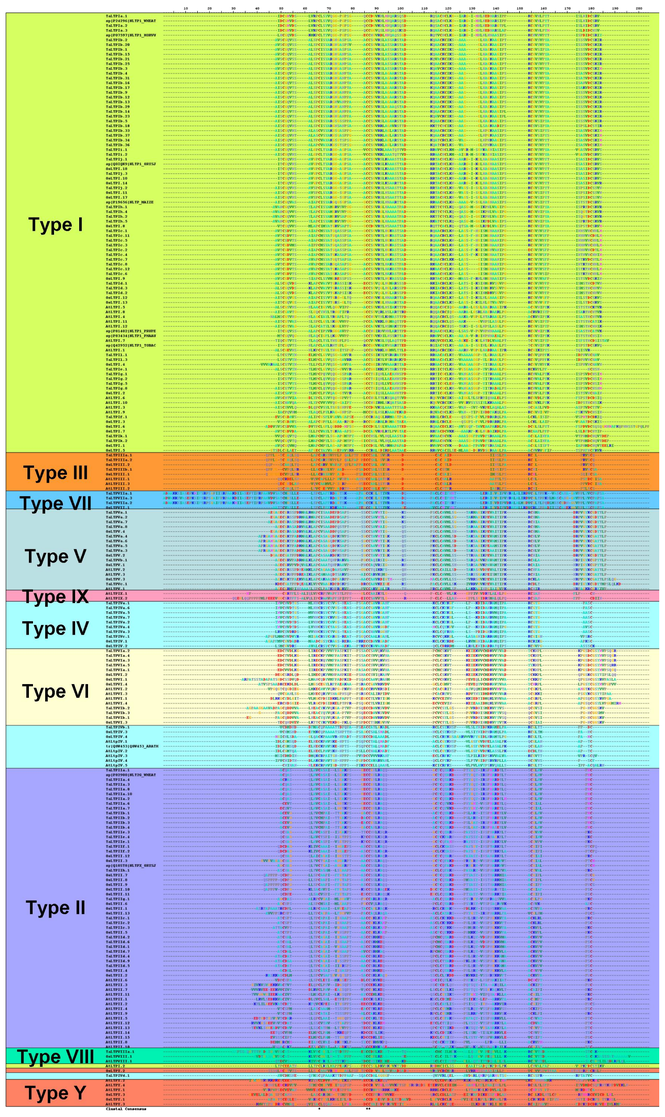


Unrooted phylogenetic tree between rice, *Arabidopsis* and wheat nsLTP gene families (Boutrot *et al.*, 2008)





Two different cysteine connectivity patterns



Multiple alignment of nsLTP mature sequences

PubMed National Library of Medicine NLM — UniProt — swissprot

sequence

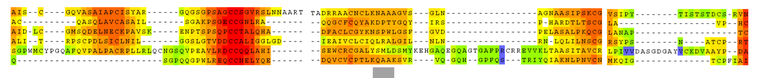
- α-helical folding pattern (4-5 helices)
- 4 disulfide bonds
- hydrophobic cavity (35-350 Å^3)
- 32 experimental structures (10 seq.)

1. target sequence(s)

```
>TaLTPIa.1
IDCGHVDSLVRPCLSYVQGGPGPSGQCCD
GVKNLHNQARSQSDRQSACNCLKGIARGI
HNLNEDNARSIPPKCGVNLPYTISLNIDC
SRV
>...
```

2. template selection

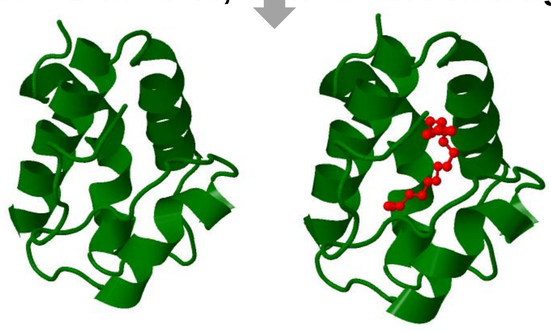
3. if multiple templates, template structure-based sequence alignment



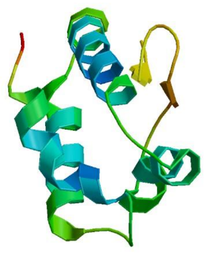
4. target-template sequence alignment

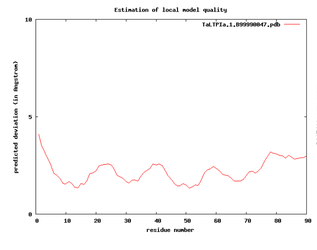
```
>P1;1fk5A
structureX:1fk5A:1    :A:93   :A:::-1.00:-1.00
AIS-C----GQVASAIAPCISYAR------GQGSGPSAGCCSGVRSLNNAARTTADRRAACNCLKNAAAGVS---GLN-------------AGNAASIPSKCGVSIPY--------TISTSTDCS-RVN*
>P1;1tukA
structureX:1tukA:1    :A:67   :A:::-1.00:-1.00
AC---------QASQLAVCASAIL-------SGAKPSGECCGNLRA---------QQGCFCQYAKDPTYGQY---IRS-------------P-HARDTLTSCGLA-----------------VP--HC*
>P1;2rknA
structureX:2rknA:1    :A:77   :A:::-1.00:-1.00
AIDLC----GMSQDELNSCRPAVSK-----ENPTSPSQPCCTALQHA--------DFACLCGYKNSPWLGSF---GVD-------------PELASALPKQCGLANAP-----------------TC*
>P1;1hypA
structureX:1hypA:6    :A:80   :A:::-1.00:-1.00
----------PSCPDLSICLNIL-------QGSLGTVDDCCALIGGLGD------IEAIVCLCIQLRALGIL---WLN-------------RN-LQLILNSCGRSYPS--------N----ATCP--RT*
>P1;1hssA
structureX:1hssA:5    :A:124  :A:::-1.00:-1.00
---MCYPQQAFQVPALPACRPLLRLQCNGSQVPEAVLRDCCQQLAHI--------SEWCRCGALYSMLDSMYKEH------GAFPRCRREVVKLTAASITAVCRLPIVVDASGDGAYVCKDVAAYP--DA*
>P1;1pnbB
structureN:1pnbB:1    :B:75   :B:::-1.00:-1.00
-----------------------------QSGPGQGGPWLRSQCCNEIYQE--------DQVCVCPTLKQAAKSVR----VQGQHGPFQS---TRIYQIAKNLPNVCNMKQIG-------------TCPFIAI*
>P1;TaLTPIa.1
sequence:TaLTPIa.1:1    : :90   : :::-1.00:-1.00
IDC------GHVDSLVRPCLSYV-------QGGPGPSGQCCDGVKNLHNQARSQSDRQSACNCLKGIARGIH---NLN-------------EDNARSIPPKCGVNLPY----------TISLNIDCSRV*
```

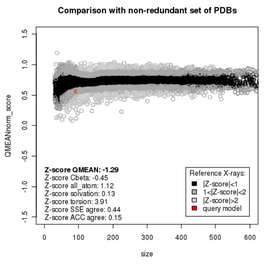
5. nsLTP 3D models, without and with ligand



modeling pipeline





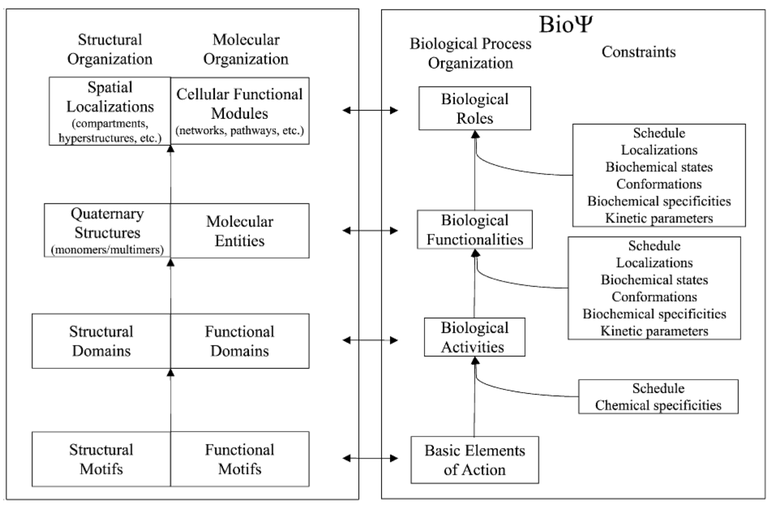


Model quality estimation

RCSB PDB PROTEIN DATA BANK — PubMed National Library of Medicine NLM — SCOP STRUCTURAL CLASSIFICATION OF PROTEINS — Pfam

structure

- > 70 ligands (lipids, hydrophobic compounds)
- roles in plant defense mechanism, resistance to abiotic stresses (desiccation, cold, salinity), germination, and others…

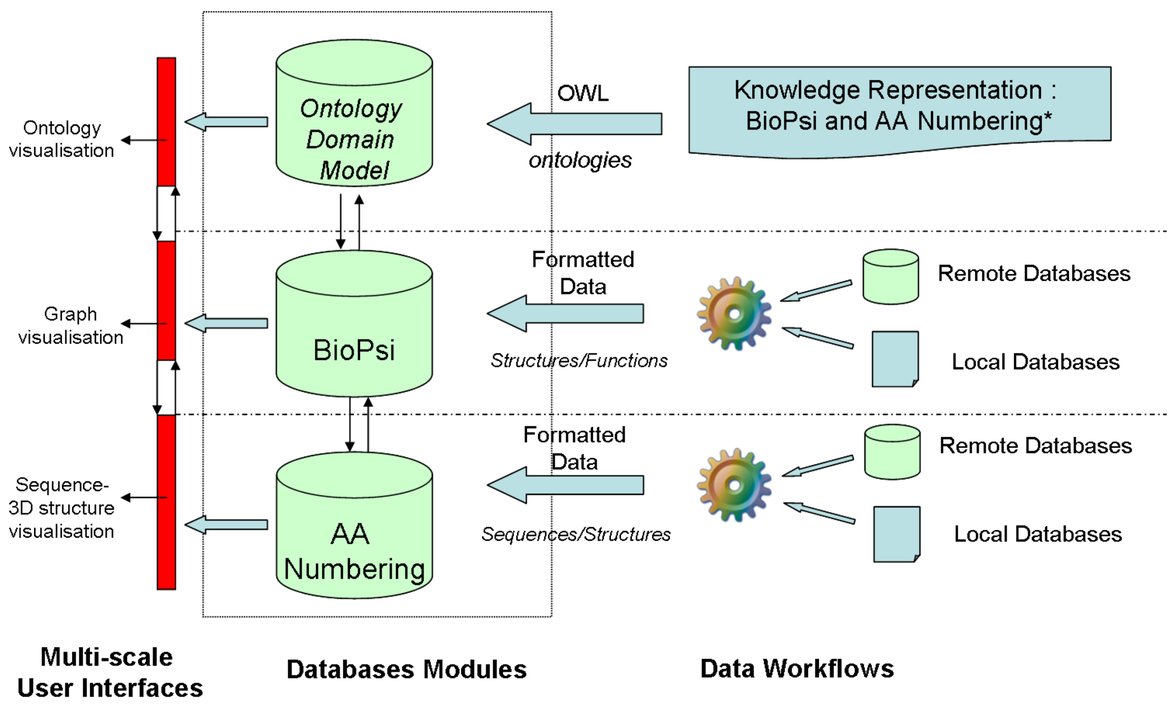


BioΨ multi-level structure–function generic description (Mazière *et al.*, 2004)

PubMed National Library of Medicine NLM

function

A multi-scale representation system



*** amino acid***

MS-DMind architecture

Agence Nationale de la Recherche ANR — South Green bioinformatics platform — cirad — SysDiag FRE3009 — INRA

[1]UMR Développement et Amélioration des Plantes, CIRAD, Montpellier, France
[2]SysDiag: Modélisation des Systèmes Complexes Biologiques pour le Diagnostic, CNRS UMR3145/Bio-Rad, Montpellier, France